LEVEL

Bolt Beranek and Newman Inc.
Report No. 2578

A.I. Report No. 13
April 1974

Selective Linear Prediction and Analysis-by-Synthesis in Speech Analysis

John Makhoul

This research was supported by: Advanced Research Projects Agency under ARPA Order No. 1967; Contract No. DAHC-71-C-0088.

$$\frac{P(\omega)}{\sigma^2} \quad \text{vs} \quad \log \frac{P(\omega)}{\sigma^2}$$

BBN Report No. 2578
A.I. Report No. 13

SELECTIVE LINEAR PREDICTION AND ANALYSIS-BY-SYNTHESIS IN SPEECH ANALYSIS

John Makhoul

April 1974

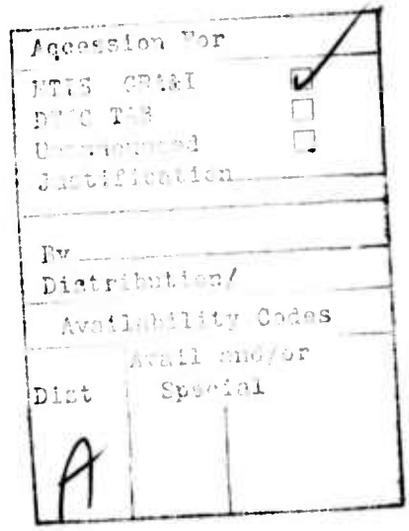
The views and conclusions contained in this document are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Defense Supply Service - Washington under Contract No. DAHC-71-C-0088. Order No. 1967.

Distribution of this document is unlimited. It may be released to the Clearinghouse, Department of Commerce for sale to the general public.

ABSTRACT

Linear prediction is presented as a spectral modeling technique in which the signal spectrum is modeled by an all-pole spectrum. The method allows for arbitrary spectral shaping in the frequency domain, and for modeling of continuous as well as discrete spectra (such as filter bank spectra). In addition, using the method of selective linear prediction, all-pole modeling is applied to selected portions of the spectrum, with applications to speech recognition and speech compression. Linear prediction is compared with traditional analysis-by-synthesis techniques for spectral modeling. It is found that linear prediction offers computational advantages over analysis-by-synthesis, as well as better modeling properties if the variations of the signal spectrum from the desired spectral model are large. For relatively smooth spectra and for filter bank spectra, analysis-by-synthesis is judged to give better results. Finally, a suboptimal solution to the problem of all-zero modeling using linear prediction is given.



ACKNOWLEDGMENTS

The author wishes to express his appreciation to R. Viswanathan for his insightful comments, and to Brenda Aighes for her invaluable help.

TABLE OF CONTENTS

I. INTRODUCTION

The short-time spectrum has been perhaps the single most important method of analysis for the study of speech. Its applications in speech synthesis, speech recognition and speaker identification are pervasive and well known. The extensive use of the short-time spectrum as an analysis tool began with the development of the sound spectrograph [1]. Even today, this three-dimensional time-frequency-intensity spectral representation is of great utility. However, there are obvious limitations on the range and flexibility of the types of analysis that can be performed, as well as limitations in the resolution and dynamic range of the output spectrogram display.

Many of the limitations of the spectrograph were overcome upon the introduction, in the 1950's, of high-speed digital computers in spectral analysis. Simultaneous with the advance in computation there were significant advances that occurred in understanding the acoustics of speech production. This was highlighted in 1960 by the publication of Fant's Acoustic Theory of Speech Production [2]. As a result of the two types of advances mentioned above, the method of spectral analysis-by-synthesis (AbS) for the reduction of speech spectra was introduced at M.I.T. and Bell Laboratories in 1961. At M.I.T. the method was used on filter-bank derived spectra to extract the

pole pattern of vowels [3,4] and pole-zero patterns of nasals [5]. At Bell Laboratories, analysis-by-synthesis was used on the computed spectrum of a single pitch period to extract the formants (resonances, poles) of the vocal tract as well as the zeros of the glottal spectrum [6].

In spectral AbS, a speech spectrum is fitted by another spectrum that is represented in terms of poles and zeros. The fit is optimized through the minimization of some error criterion. The error between the two log spectra is minimized in an iterative manner. The early attempts minimized the error by recursively varying only one pole or zero at a time. These methods were error prone and were not easily adaptable to an automatic algorithm. More recently, Olive [7] developed a Newton-Raphson technique that performs the iterative computation on all poles simultaneously and in a straightforward automatic manner.

In this paper we present another method of spectral modeling which makes use of recent advances in the field of digital signal processing, in particular the introduction of linear prediction (LP) to speech analysis. The major difference between AbS and LP analysis is the error criterion used in the matching process, which in the latter is the integrated ratio of the two spectra. In general, this error criterion leads to a better spectral envelope fit. In addition, for the special (but important) case

of an all-pole model spectrum, LP analysis offers two important advantages: (a) The computations for the spectral parameters are straightforward and noniterative, and (b) if the time signal is available there is no need to compute the spectrum first. The two methods have two properties in common: (a) The spectral matching can be done selectively to any portion of the spectrum, and (b) both error criteria are functions of the ratio of the original and model spectra, thereby resulting in a matching process that is uniform over the frequency range of interest.

In Section II we apply LP analysis to spectral matching by all-pole model spectra. Section III describes the properties of the optimum model spectrum. In Section IV we introduce the method of selective linear prediction, where LP analysis is applied to a selected portion of the spectrum, and we describe its applications to speech recognition and speech compression. Section V describes the application of LP analysis to the modeling of discrete spectra (such as harmonic spectra and those obtained from a bank of filters). Section VI examines the properties of the error measure used in LP analysis and gives a critical comparison between LP analysis and analysis-by-synthesis. Section VII gives a suboptimal solution to the problem of all-zero modeling using LP analysis.

II. LINEAR PREDICTIVE SPECTRAL MODELING

Let us assume that we are given a power spectrum $P(\omega)$ with bandwidth B, i.e. $P(\omega)$ is known for $0 \leq \omega \leq \omega_b = 2\pi B$. (The more general case where the frequency range covers only a portion of a spectrum is treated later.) In this method, we shall view $P(\omega)$ as the spectrum of some signal $s(nT)$ that was sharply low-pass filtered at B Hz and sampled at a frequency $f_s = 2B = \frac{1}{T}$, where T is the sampling period. We shall view $P(\omega)$ as such irrespective of how it was actually generated. This now allows us to deal with $P(\omega)$ as the spectrum of a sampled signal and, hence, we can make use of digital signal processing techniques. In particular, instead of using the complex s plane we now use the complex z plane. In essence, we map $P(\omega)$ onto the upper half of the unit circle in the z plane such that the angular distance $\theta = \omega T$. The mapping is such that $\omega = 0$ corresponds to $\theta = 0$ and $\omega = \omega_b$ corresponds to $\theta = \pi$. In addition $P(-\omega) = P(\omega)$ defines the spectrum over the bottom half of the unit circle, i.e. the spectrum is even and real. (For convenience, we shall set the sampling interval T=1. For other values of T simply replace ω by ωT in the appropriate equations.) Thus, we shall assume that

$$P(\omega) = |S(e^{j\omega})|^2 \quad , \tag{1}$$

where $S(z)$ is the z transform of the hypothetical signal s_n.

We wish to fit $P(\omega)$ in some optimal manner by an all-pole spectrum $\hat{P}(\omega)$. Let us assume that the model spectrum corresponds to a transfer function $\hat{S}(z)$ given by

$$\hat{S}(z) = \frac{G}{A(z)} = \frac{G}{1 + \sum_{k=1}^{p} a_k z^{-k}} , \tag{2}$$

where

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \tag{3}$$

will be called the inverse filter, p is the number of poles in the model spectrum, and G is a constant gain factor. The model spectrum $\hat{P}(\omega)$ is then given by

$$\hat{P}(\omega) = |\hat{S}(e^{j\omega})|^2 = \frac{G^2}{|A(e^{j\omega})|^2}$$

$$= \frac{G^2}{\left|1 + \sum_{k=1}^{p} a_k e^{-jk\omega}\right|^2} . \tag{4}$$

Given a spectrum $P(\omega)$ and a number of poles p, we must determine the parameters $\{a_k, 1 \le k \le p\}$ and G.

We define an error measure E between $P(\omega)$ and $\hat{P}(\omega)$:

$$E = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)} \, d\omega \tag{5}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) \, |A(e^{j\omega})|^2 \, d\omega . \tag{6}$$

E can be interpreted as the total energy of the "error signal" obtained by passing the hypothetical signal s_n through the inverse filter A(z). (This is clear by using Parseval's theorem.) Note from (6) that E is defined to be independent of G. The gain factor is determined from energy considerations.

The parameters $\{a_k\}$ are determined by minimizing E in (6) with respect to each of the parameters. This is accomplished by setting

$$\frac{\partial E}{\partial a_i} = 0, \quad 1 \le i \le p \ . \tag{7}$$

From (4-6) it can be shown that [8]

$$\frac{\partial E}{\partial a_i} = 2\left[R_i + \sum_{k=1}^{p} a_k \, R_{|i-k|}\right] , \tag{8}$$

where

$$R_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega) \cos(k\omega) \, d\omega \tag{9}$$

is the autocorrelation function corresponding to the signal spectrum $P(\omega)$. From (7) and (8) we must have

$$\sum_{k=1}^{p} a_k \, R_{|i-k|} = -R_i, \quad 1 \le i \le p \ . \tag{10}$$

This is a set of p linear equations in p unknowns which can be solved for the parameters $\{a_k\}$ of the all-pole model spectrum.

A recursive solution is given elsewhere [8,15,16].

The minimum error is obtained by substituting (9) and (10) in (6). The result can be shown to be

$$E_p = R_o + \sum_{k=1}^{p} a_k R_k \ , \tag{11}$$

where the dependence of the minimum error on p is shown explicitly.

The gain factor G^2 in (4) is obtained by conserving energy between the original and model spectra, i.e.

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{P}(\omega)\, d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} P(\omega)\, d\omega$$

or
$$\hat{R}_o = R_o \ , \tag{12}$$

where
$$\hat{R}_i = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{P}(\omega) \cos(i\omega)\, d\omega \tag{13}$$

is the autocorrelation function corresponding to the model spectrum. An analytic expression for $\hat{R}_i$ is more easily obtained from the unit sample response $\hat{s}_n$ of $\hat{S}(z)$ in (2). Taking the inverse z transform of (2) we have

$$\hat{s}_n = \begin{cases} 0 \ , & n<0, \\ G \ , & n=0, \\ -\sum_{k=1}^{p} a_k \hat{s}_{n-k} \ , & n>0. \end{cases} \tag{14}$$

By definition, the autocorrelation function $\hat{R}_i$ is given by

$$\hat{R}_i = \sum_{n=-\infty}^{\infty} \hat{s}_n \hat{s}_{n+|i|} \quad . \tag{15}$$

From (14) and (15) it can be shown that

$$\hat{R}_o = G^2 - \sum_{k=1}^{p} a_k \hat{R}_k \tag{16}$$

and

$$\hat{R}_i = -\sum_{k=1}^{p} a_k \hat{R}_{|i-k|} \quad , \quad 1 \leq |i| \leq \infty \quad . \tag{17}$$

From (10), (12), (16) and (17), we conclude that

$$\hat{R}_i = R_i \quad , \quad 0 \leq i \leq p \quad , \tag{18}$$

and

$$G^2 = R_o + \sum_{k=1}^{p} a_k R_k \quad . \tag{19}$$

Therefore, from (11) and (19), G^2 is equal to the minimum error E_p.

Equations (10) and (19) completely specify the model spectrum $\hat{P}(\omega)$. Given a spectrum $P(\omega)$ and a desired number of poles p, the parameters of $\hat{P}(\omega)$ are obtained by first computing the autocorrelation coefficients R_i, $0 \leq i \leq p$, using (9). The coefficients $\{a_k\}$ are then computed from (10) and the gain G from (19).

Equivalently, if the speech signal itself is given, it is not necessary to compute $P(\omega)$ first. Instead, the autocorrelation

coefficients R_i can be computed from the signal directly:

$$R_i = \sum_{n=-\infty}^{\infty} s_n s_{n+|i|} , \quad 0 \le i \le p \quad . \tag{20}$$

It is clear that (20) can be evaluated only if the signal is of finite duration. This usually brings up the issue of windowing. (See Makhoul and Wolf [8] for a discussion of windowing of speech signals.)

The spectral fitting method described in this section can be shown to be equivalent to the autocorrelation method of linear prediction [8,9], where the coefficients a_k are the predictor coefficients. That is why we have chosen to call this method the linear predictive (LP) spectral modeling method. The model spectrum $\hat{P}(\omega)$ is also known as the LP spectrum.

Figure 1 shows an example of LP spectral matching for a spectrum over 0-10 kHz with the number of poles p=28. In this case the original spectrum $P(\omega)$ was obtained by computing the fast Fourier transform (FFT) of a 20 ms, Hamming windowed, 20 kHz sampled speech signal. The spectrum $\hat{P}(\omega)$ was computed from (4) by dividing G^2 by the magnitude squared of the FFT of the sequence; $1, a_1, a_2, \ldots, a_p$. Arbitrary frequency resolution can be obtained by simply appending an appropriate number of zeros to this sequence before taking the FFT.

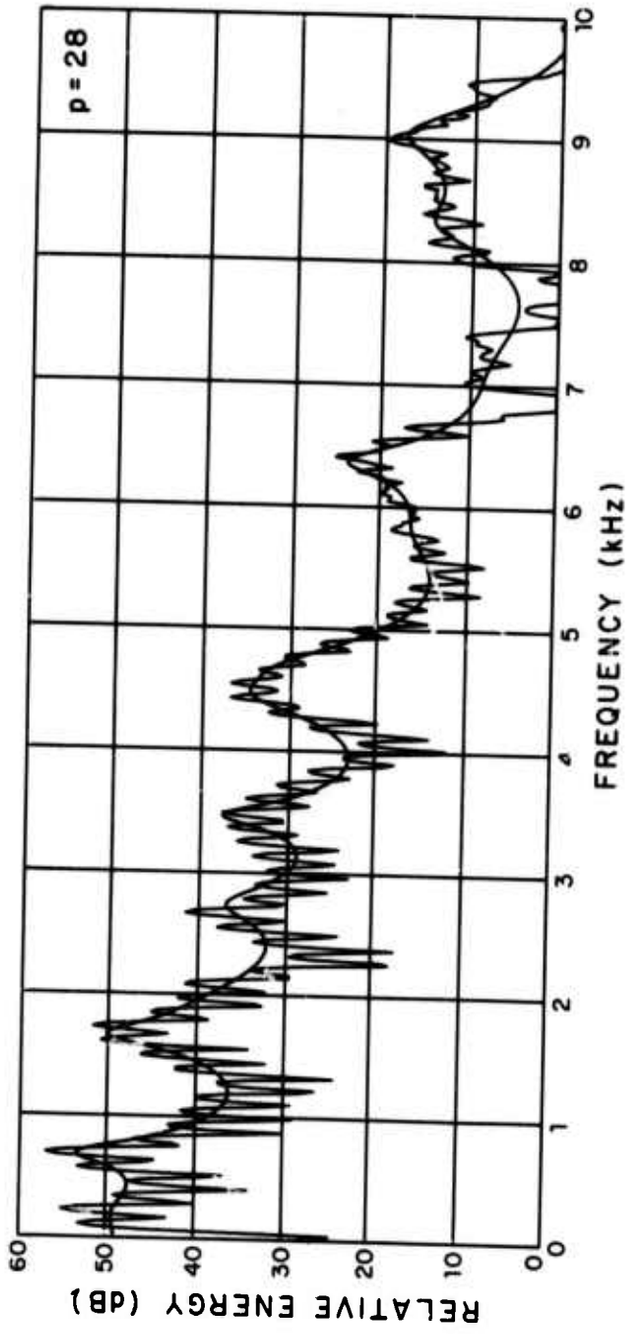


Fig. 1. A 28-pole spectral fit to an FFT-computed signal spectrum.

Before we turn to the general case of selective spectral matching, we shall examine the properties of the model spectrum $\hat{P}(\omega)$.

III. PROPERTIES OF THE MODEL SPECTRUM

The poles of the model spectrum can be found by computing the roots of the polynomial A(z) in (3). Since the coefficients a_k are real, some or none of the roots are real and the rest are complex conjugate pairs. Conversion of the poles to the s plane can be achieved by setting each root $z_k = e^{s_k T}$, where $s_k = \sigma_k + j\omega_k$ is the corresponding pole in the s plane. If the root $z_k = z_{kr} + jz_{ki}$, then:

$$\omega_k = \frac{1}{T} \arctan \frac{z_{ki}}{z_{kr}} \quad , \tag{21a}$$

$$\sigma_k = \frac{1}{2T} \log(z_{kr}^2 + z_{ki}^2) \quad , \tag{21b}$$

where z_{kr} and z_{ki} are the real and imaginary parts of z_k, respectively, and T is the sampling period.

One important property of the poles of $\hat{S}(z)$ is that they are guaranteed to be inside the unit circle, provided $P(\omega)$ is a positive definite spectrum [10].

For a well chosen number of poles p, some of the poles of $\hat{S}(z)$ can be related to vocal tract resonances. The extent to which the formant values thus obtained reflect the actual resonances of the vocal tract depends on several factors, including the adequacy of the all-pole model for each spectrum considered, and the number of poles in the model. These issues are

discussed in more detail elsewhere [8].

The manner in which the model spectrum $\hat{P}(\omega)$ approximates $P(\omega)$ is reflected largely in (18), which relates the autocorrelation coefficients $\hat{R}_i$ and R_i. Since $P(\omega)$ and $\hat{P}(\omega)$ are the Fourier transforms of R_i and $\hat{R}_i$, respectively, it follows that increasing the value of p increases the range over which R_i and $\hat{R}_i$ are equal, resulting in a better fit of $\hat{P}(\omega)$ to $P(\omega)$. In the limit, as $p \to \infty$, $\hat{R}_i$ becomes identical to R_i for all i, and hence the two power spectra become identical:

$$\hat{P}(\omega) = P(\omega) \text{ , as } p \to \infty \quad . \tag{22}$$

Since the minimum error $E_p = G^2$, we have from (5):

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P(\omega)}{\hat{P}(\omega)} \, d\omega = 1 \quad . \tag{23}$$

Equation (23) is true for all values of p. In particular, it is true as $p \to \infty$, in which case from (22) we see that (23) becomes an identity. Another important case where (23) becomes an identity is when $P(\omega)$ is an all-pole spectrum with p_0 poles, then $\hat{P}(\omega)$ will be identical to $P(\omega)$ for all $p \geq p_0$. Relation (23) will be useful in discussing the properties of the error measure in Section VI.

Another property of $\hat{P}(\omega)$ that will be discussed later on is that the slope of $\hat{P}(\omega)$ is zero at $\omega=0$ and $\omega=\pi$:

$$\frac{\partial \hat{P}(\omega)}{\partial \omega} = 0 \quad , \quad \omega=0,\pi \quad . \tag{24}$$

This can be easily seen by rewriting (4) as

$$\hat{P}(\omega) = \frac{G^2}{b_0+2\sum_{k=1}^{p} b_k \cos(k\omega)} \quad , \tag{25}$$

where

$$b_k = \sum_{n=0}^{p-|k|} a_n a_{n+|k|} \quad , \; a_o = 1, \quad 0\le k\le p \quad , \tag{26}$$

are the autocorrelation coefficients of the impulse response of the inverse filter $A(z)$. By taking $\frac{\partial \hat{P}(\omega)}{\partial \omega}$ in (25), it is clear that it is equal to zero at 0 and π.

Equation (25) gives another method for computing $\hat{P}(\omega)$, and that is by dividing G^2 by the real part of the FFT of the sequence: $b_0, 2b_1, 2b_2, \ldots, 2b_p$.

IV. SELECTIVE LINEAR PREDICTION

We now generalize the LP spectral modeling method to the case where we wish to fit a selected portion of a given spectrum.

In general, we have a spectrum $P(\omega)$, $0 \le \omega \le \omega_b$, and we wish to match the spectrum in a region (Ω: $\omega_\alpha \le \omega \le \omega_\beta$) by an all-pole spectrum $\hat{P}(\omega)$ as given by (4). Call the spectrum in the region Ω, $P'(\omega)$. In order to compute the parameters of $\hat{P}(\omega)$ we simply map the region Ω onto the unit circle such that $\omega_\alpha \to 0$ (the arrow is read "mapped into") and $\omega_\beta \to \pi$, and then follow the same procedure outlined in Section II. The mapping is done as follows:

$$\text{Define} \qquad \omega' = \omega - \omega_\alpha$$

$$\text{and} \qquad \omega_b' = \omega_\beta - \omega_\alpha \ .$$

$$\text{Then} \qquad \Omega \to 0 \le \omega' \le \omega_b' \qquad (27)$$

$$\text{and} \qquad T' = \frac{\pi}{\omega_b'} \ ,$$

where T' is a new hypothetical sampling interval. The problem now reduces to the original one. The autocorrelation coefficients R_k, $0 \le k \le p$, are computed from (9) with $P'(\omega')$ replacing $P(\omega)$. Then (10) and (19) are used to solve for the parameters of the model spectrum $\hat{P}(\omega')$.

Figure 2 shows an example of selective linear prediction. The signal spectrum is identical to that shown in Figure 1. In Figure 2 the two halves of the spectrum were matched separately by a 14-pole model spectrum in each half. In the left half ω_α=0 and ω_β=5 kHz, and in the right half ω_α=5 kHz and ω_β=10 kHz. Since the matching for each half was done independently, there is no guarantee that the two model spectra will join smoothly at 5 kHz. In fact, in general, a discontinuity such as the one in Figure 2 is expected. Recall that the model spectrum has zero slope at 0 and π. This is evident in Figure 2 at 5 kHz. The reader will also note other differences between Figs. 1 and 2 in the manner in which the original and model spectra match.

Figure 3 shows the same signal spectrum as in Figure 2, but with the right half of the spectrum being fitted by only a 5-pole spectrum. This demonstrates the flexibility of selective linear prediction in that different portions of a spectrum can be matched using different numbers of poles.

Applications to Speech Recognition and Compression

Here we shall demonstrate the idea of selective linear prediction as applied to speech recognition and speech compression. It is important to note that, since we assume the availability of the signal spectrum $P(\omega)$, any desired frequency shaping or

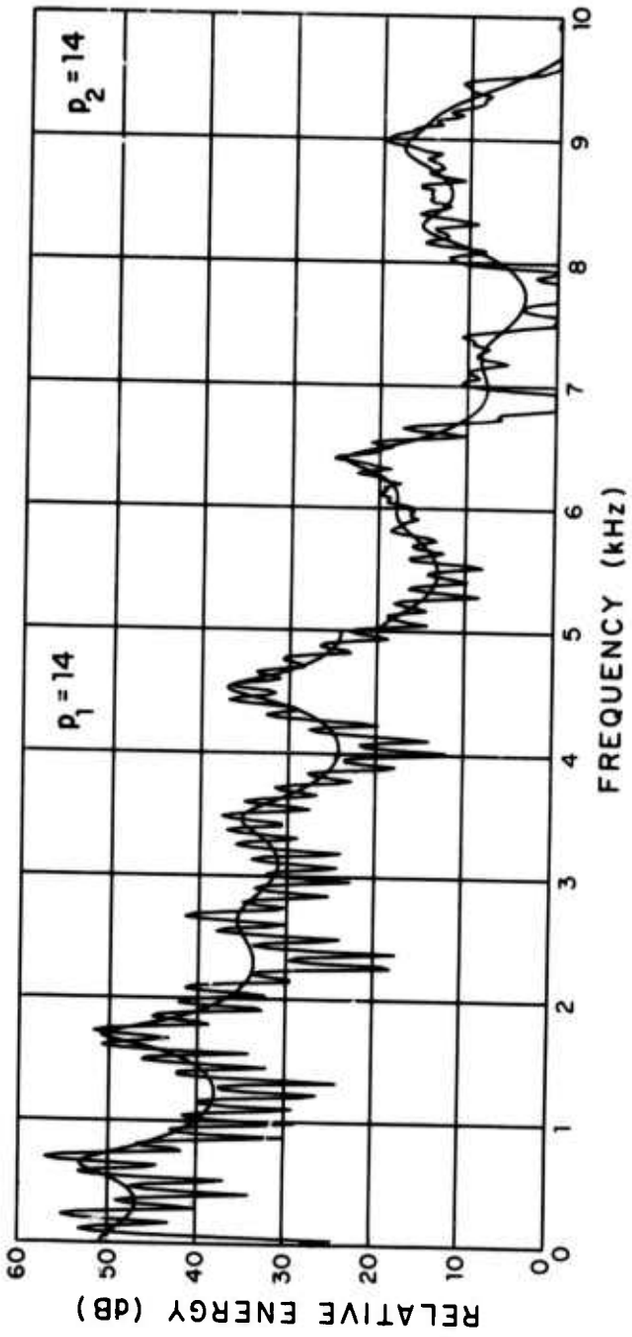


Fig. 2. Application of selective linear prediction to the same signal spectrum as in Fig. 1, with independent 14-pole model fits to the 0-5 and 5-10 kHz regions.

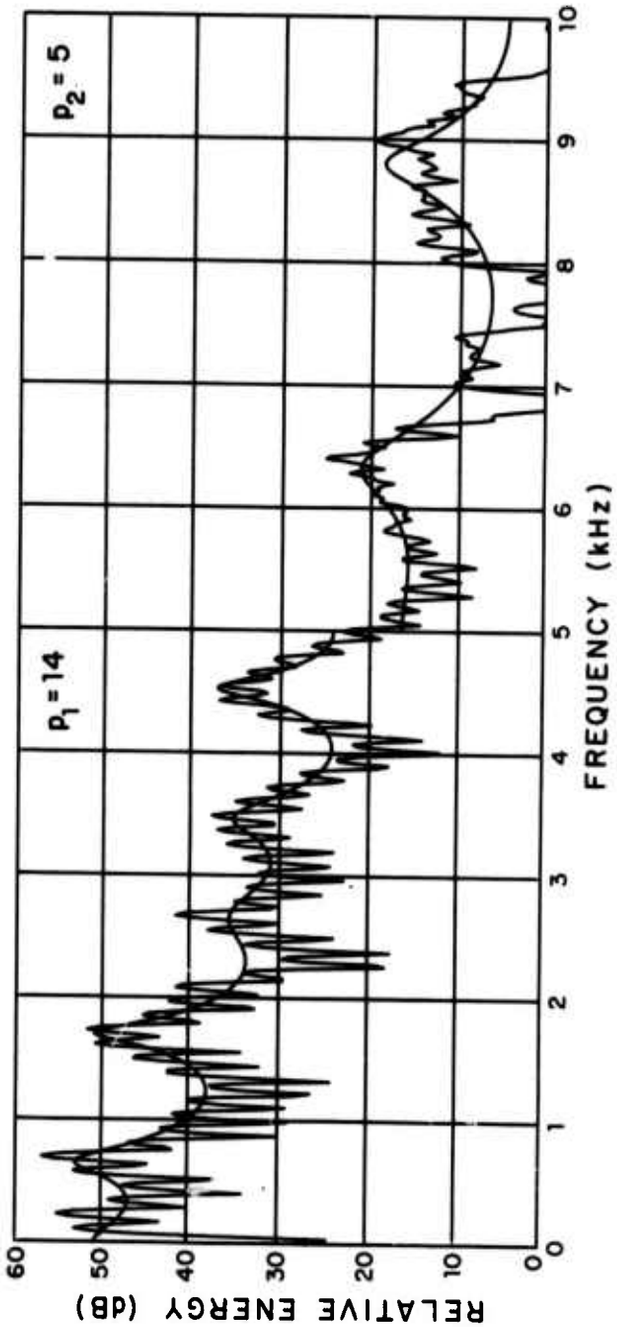


Fig. 3. Application of selective linear prediction to the same signal spectrum as in Fig. 1, with a 14-pole fit to the 0-5 kHz region and a 5-pole fit to the 5-10 kHz region.

filtering can be done directly to the signal spectrum before LP analysis is performed. This will be clear below.

Figures 1-3 show the same signal spectrum computed from a 20 kHz sampled signal. Now, in order to model the spectral envelope for the whole frequency range from 0-10 kHz, one would probably use anywhere between 24-28 poles for the all-pole model spectrum. A 28-pole fit is shown in Fig. 1. For speech recognition applications, however, the main region of interest is the 0-5 kHz region. The spectrum in the 5-10 kHz region is of interest mainly for the recognition of fricatives, in which case the total energy in that region might be sufficient. We also know that in LP analysis the spectral matching process performs uniformly over the whole frequency range, which might not be desirable in this case because the all-pole assumption for many speech sounds is less applicable for frequencies greater than 5 kHz. Therefore, instead of modeling the whole spectrum, we use selective LP to model the lower 5 kHz by a lower order all-pole spectrum. A 14-pole fit is shown in Figs. 2 and 3. In this manner, not only do we reduce our computations for the poles, but we are also in the advantageous position of having to interpret 14 instead of 28 poles. The total energy in the 5-10 kHz region can be easily computed directly from the spectrum and used for the detection of fricatives if desired. Alternatively one could

fit a very low order all-pole spectrum to that region, as shown in Fig. 3.

Now, the same type of analysis could have been done in the time domain, but consider what one would have had to do. First, the 20 kHz sampled signal must be sharply filtered at 5 kHz. Second, and very importantly, the signal must be down-sampled to 10 kHz by discarding every other sample. Third, a 14-pole LP analysis is performed on the resulting signal. And fourth, in order to obtain the energy in the 5-10 kHz region, one subtracts the energy in the 10 kHz signal from the energy in the original 20 kHz signal. (It is even more complicated if one wants to perform an LP analysis on the 5-10 kHz region in the time domain.)

Not only is the time domain analysis more involved and costly; it is also very inflexible. Consider the problem of having to carry the same procedure to match the spectrum in the 0-3.5 kHz region instead of 0-5 kHz. In that case, it would be necessary to perform the time-domain down-sampling from 20 kHz to 7 kHz: a rather difficult task. The elegance of the method of selective linear prediction lies in the fact that the two problems of sharp filtering and down sampling are completely solved by working in the frequency domain.

We are currently applying this property to speech compression systems that employ linear prediction. In this application, it is desirable to be able to test the performance of the system at different sampling rates. We sample the signal at the highest sampling rate desired, and then we simulate the performance of different sampling rates by applying selective linear prediction to the corresponding frequency bands.

V. MODELING DISCRETE SPECTRA

Thus far we have assumed that the spectrum $P(\omega)$ is a continuous function of frequency. Most often, however, the spectrum is known at only a finite number of frequencies. For example, an FFT-derived spectrum has values at equally spaced frequency points. On the other hand, filter bank spectra usually have values at frequencies that are not necessarily equally spaced. For these discrete cases we define the error measure E as a summation instead of an integral:

$$E = \frac{G^2}{N} \sum_{n=0}^{N-1} \frac{P(\omega_n)}{\hat{P}(\omega_n)} \quad , \tag{28}$$

where N is the total number of spectral points on the unit circle. Following the same minimization procedure as in the continuous case, we obtain the set of equations (10) again, but the coefficients R_k are now defined as

$$R_k = \frac{1}{N} \sum_{n=0}^{N-1} P(\omega_n) \cos(k\omega_n) \quad . \tag{29}$$

Note that in (28), only values of $\hat{P}(\omega)$ at the frequencies ω_n contribute to the total error. Therefore, after $\hat{P}(\omega)$ is obtained, the error between $P(\omega)$ and $\hat{P}(\omega)$ is minimum at the frequencies ω_n , $0 \le n \le N-1$. At other frequencies, $\hat{P}(\omega)$ cannot be guaranteed

in any way except in that it is a smooth function of frequency as given by (4).

If the spectrum is known at equally spaced frequency points, then if desired, (29) can be computed via a fast Fourier transform (FFT) of the spectrum $P(\omega_n)$. (In that case a highly composite value of N would help.) However, if the spectrum $P(\omega_n)$ is known at frequencies that are not equally spaced, then one can define a new spectrum $Q(\omega_m)$ at equally spaced frequencies such that $Q(\omega_m) = P(\omega_n)$ at every ω_n, and is zero otherwise. One can then use an FFT on $Q(\omega_m)$ to compute R_k. We do not necessarily recommend the use of the method just outlined for cases where the frequency spacing is nonuniform, because very often it is simply faster to compute (29) directly. However, we wished to make the point that adding spectral values that are zero does not affect the error minimization process in any way, since those values do not contribute to the total error, as is clear from (28).

Computational Considerations

The solution for the predictor coefficients a_k in (10) is unaffected if each of the autocorrelation coefficients is multiplied or divided by a constant. Therefore, the division by N in (29) is unnecessary to obtain the desired solution of (10).

The only possible importance of the division by N (or some other number) is to get a good estimate of the total energy R_o. What number to divide with depends on how the signal spectrum was obtained and on the particular application.

The spectrum $P(\omega_n)$ is an even function of frequency, i.e. $P(\omega_{N-n}) = P(\omega_n)$. Usually what we have is a spectrum that we map onto the unit circle, as explained in Section IV. The evenness property is then applied in order to complete the definition of the spectrum around the unit circle. The mapping in the continuous frequency case is no problem. However, there are a few matters to worry about in the discrete case. The main problem is the relation of the frequencies ω_α and ω_β in (27) to the discrete frequencies ω_n. There is a total of four possible cases which are divided in two categories:

(a) N even

(1) $\omega_o \to 0$, $\omega_{N/2} \to \pi$.

(2) None of the frequencies ω_n correspond to either 0 or π.

(b) N odd

(1) $\omega_o \to 0$.

(2) $\omega_{\frac{N-1}{2}} \to \pi$.

The four cases are illustrated in Fig. 4, where the crosses on the unit circle correspond to the frequencies ω_n. Case (a1) is the one usually encountered in FFT-derived spectra with even N. Case (a2) is usually encountered with filter bank spectra. Note that, because of the evenness property of $P(\omega_n)$, (29) can be simplified, but in a slightly different manner for each of the four cases.

Application to Filter Bank Spectra

We simulated the output of a filter bank by simply adding the energy in specified frequency bands from an FFT-derived spectrum. The resulting simulated filter bank has center frequencies and bandwidths similar to the hardware filter bank at the Speech Communication Laboratory at M.I.T. The filters are linearly spaced up to 1.6 kHz and logarithmically spaced thereafter. Figures 5 and 6 show two examples of the application of LP spectral modeling to the outputs of the simulated filter bank. In each figure, the original spectrum and the corresponding simulated filter bank spectrum are shown along with a 14-pole LP spectrum in each case. (The spectral lines in the filter bank spectra are shown with a finite width only because of the manner in which they were plotted.) The filter bank LP spectra in Figs. 5b and 6b are quite similar to those in Figs. 5a and 6a,

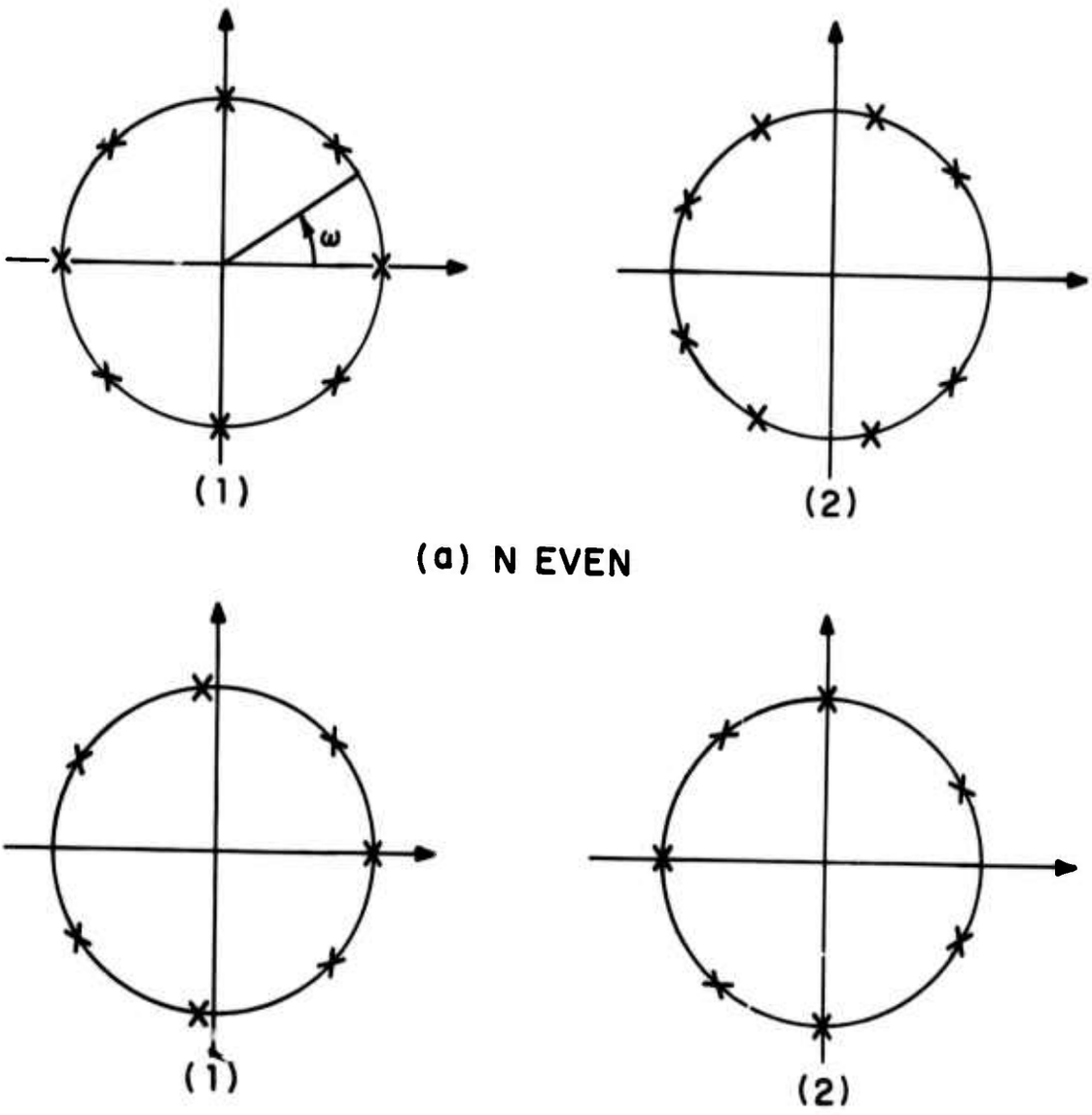


Fig. 4. Four possible configurations for discrete spectra. Each cross represents one of the N spectral lines in the spectrum.

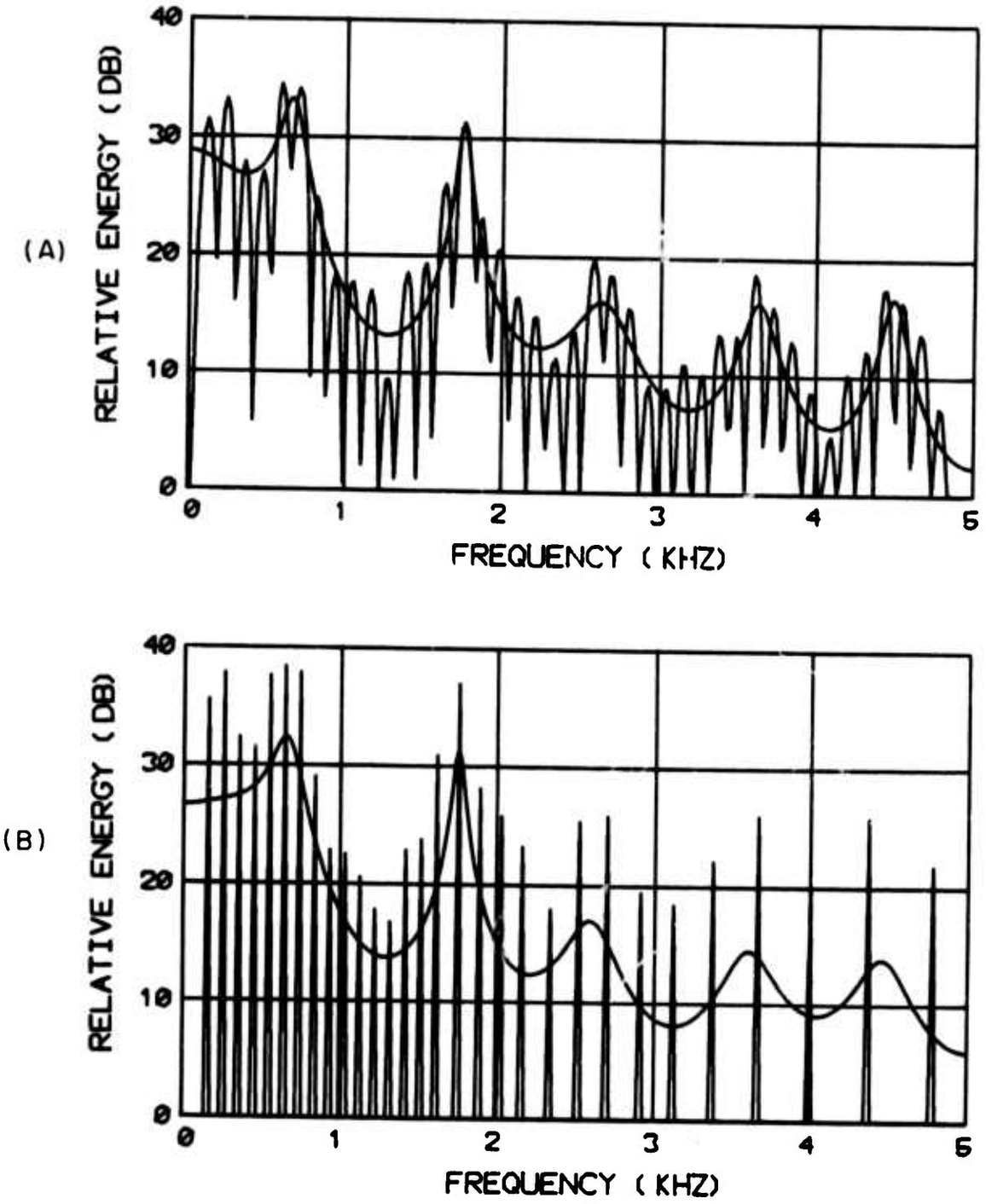


Fig. 5. Application of LP modeling to a filter bank vowel spectrum.
(A) A 14-pole fit to the original spectrum.
(B) A 14-pole fit to the simulated filter bank spectrum.

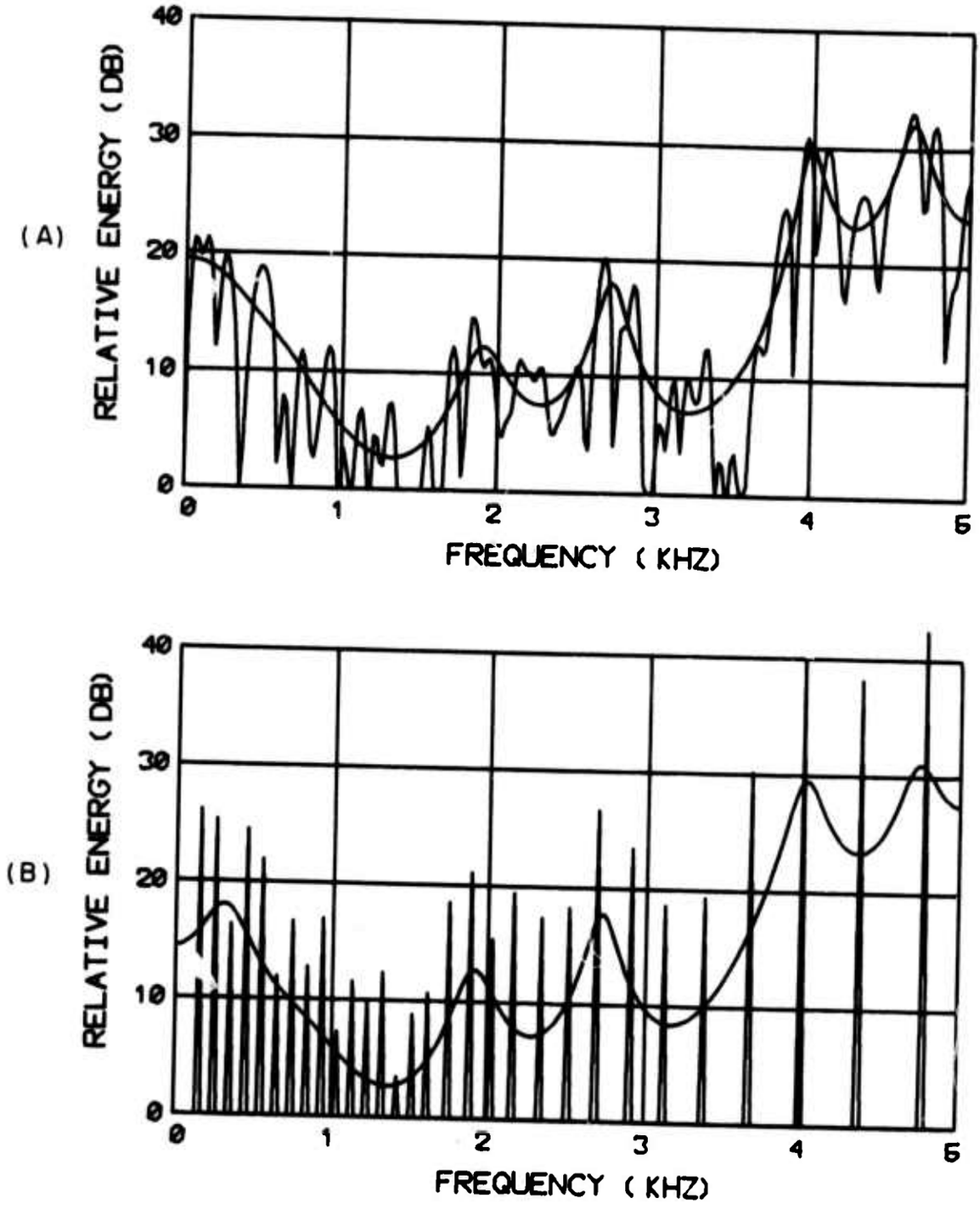


Fig. 6. Application of LP modeling to a filter bank fricative spectrum. (A) A 14-pole fit to the original spectrum. (B) A 14-pole fit to the simulated filter bank spectrum.

in spite of the relatively few spectral points in the filter bank spectra, especially at high frequencies. The extra peak at low frequencies in Fig. 6b is due to the lack of spectral points at frequencies less than 150 Hz.

Spectra of Periodic Signals

We have seen in Section III that if the signal spectrum $P(\omega)$ consists of p_o poles only, then for $p=p_o$ the LP spectrum $\hat{P}(\omega)$ is identical to $P(\omega)$. The situation is not so favorable for discrete signal spectra, as we shall see below.

Let us assume that we are given a discrete spectrum $P_1(\omega)$ that has values at equally spaced frequencies with a spacing of ω_o, such that

$$P_1(\omega) = \begin{cases} P_o(\omega) , & \omega=n\omega_o , \text{ n integer} , \\ 0 , & \text{otherwise,} \end{cases} \tag{30}$$

where $P_o(\omega)$ is a p_o-pole spectrum. $P_1(\omega)$ can be regarded as the spectrum of a periodic signal that is generated by applying a periodic unit sample sequence with period $\tau = \frac{2\pi}{\omega_o}$ to an all-pole filter whose magnitude squared frequency response is given by $P_o(\omega)$. The question is, if $P_1(\omega)$ is our signal spectrum, what will be the corresponding LP model spectrum for $p=p_o$? For LP modeling in the discrete case we compute the parameters a_k from

(10), where the autocorrelation coefficients R_k are computed from the DFT in (29) with $P(\omega_n)$ replaced by $P_1(n\omega_0)$. For a nonzero fundamental ω_0, the resulting model spectrum $\hat{P}_1(\omega)$ will <u>not</u> be equal to $P_0(\omega)$ for $p=p_0$, or any other value of p. This is illustrated in Fig. 7a where $P_0(\omega)$ is the dashed curve, $P_1(\omega)$ is the line spectrum with $F_0 = \frac{\omega_0}{2\pi} = 312$ Hz, and $\hat{P}_1(\omega)$ is the solid curve and represents the LP spectrum corresponding to $P_1(\omega)$ for $p=p_0$ (here $p_0=14$). The discrepancy between $\hat{P}_1(\omega)$ and $P_0(\omega)$ in Fig. 7a is obvious. A decrease in F_0 brings $\hat{P}_1(\omega)$ closer to $P_0(\omega)$ as in Fig. 7b. In the limit as F_0 approaches zero ($\omega_0 \to 0$), $P_1(\omega)$ approaches $P_0(\omega)$ and $\hat{P}_1(\omega)$ becomes identical to $P_0(\omega)$, as we already know from the continuous frequency case.

Figures 8 and 9 show other examples of modeling spectra of periodic signals. The types of discrepancies that can occur between the model and original spectra include merging or splitting of pole peaks, and increasing or decreasing of pole frequencies and bandwidths. In general, the pole movements are in the direction of the nearest harmonic. Atal [11] has been making quantitative measurements of these discrepancies.

It is important to note in Figs. 7-9 that the dashed curve in each case is the <u>only</u> possible p_0-pole spectrum that coincides with the line spectrum at the harmonics. (In general this is

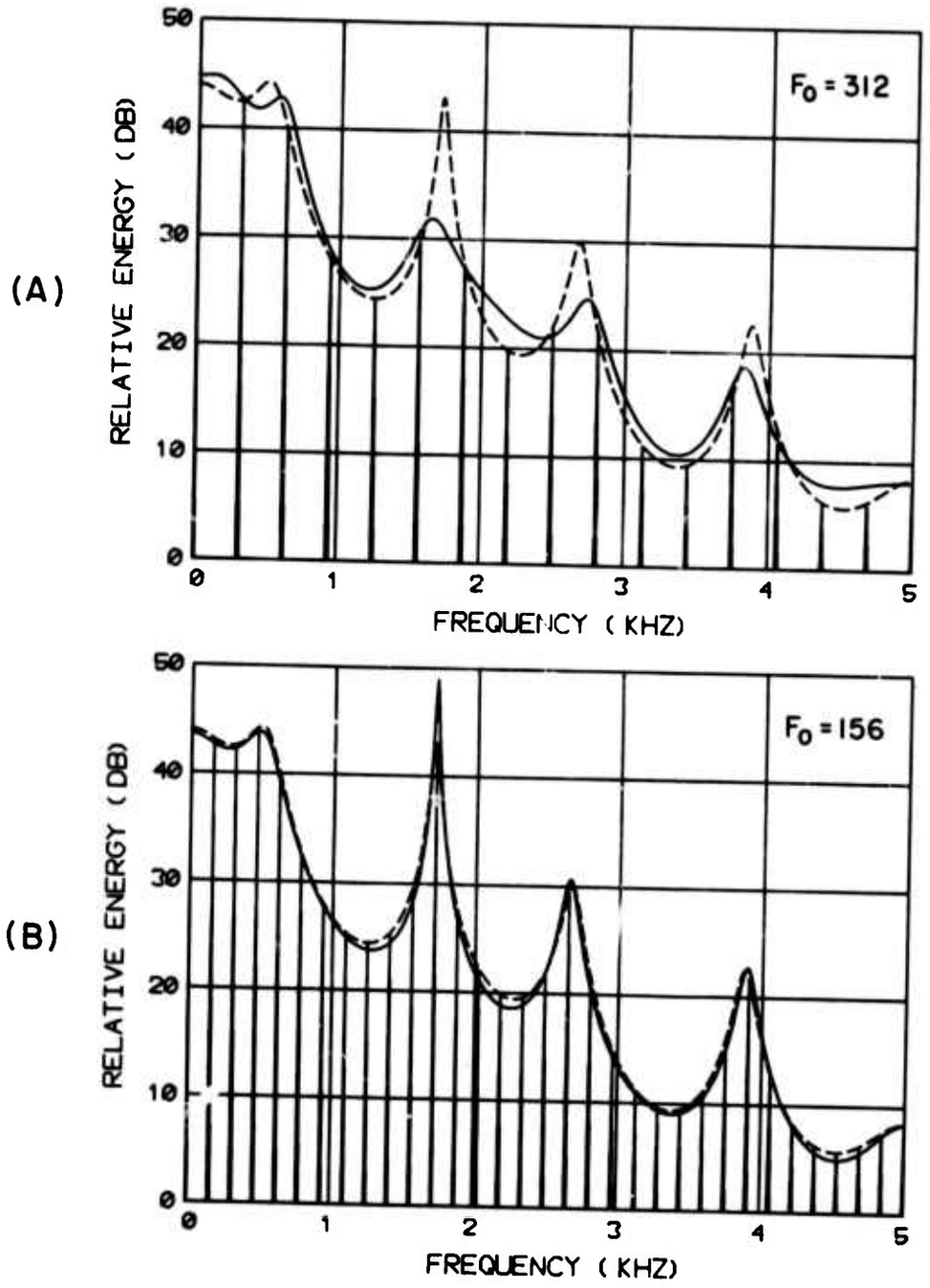


Fig. 7. LP modeling of harmonic spectra.
Dashed curve: Filter 14-pole spectrum.
Vertical lines: Corresponding harmonic spectrum for (A) F_o=312 Hz, and (B) F_o=156 Hz.
Solid curve: A 14-pole fit to the discrete harmonic spectrum.

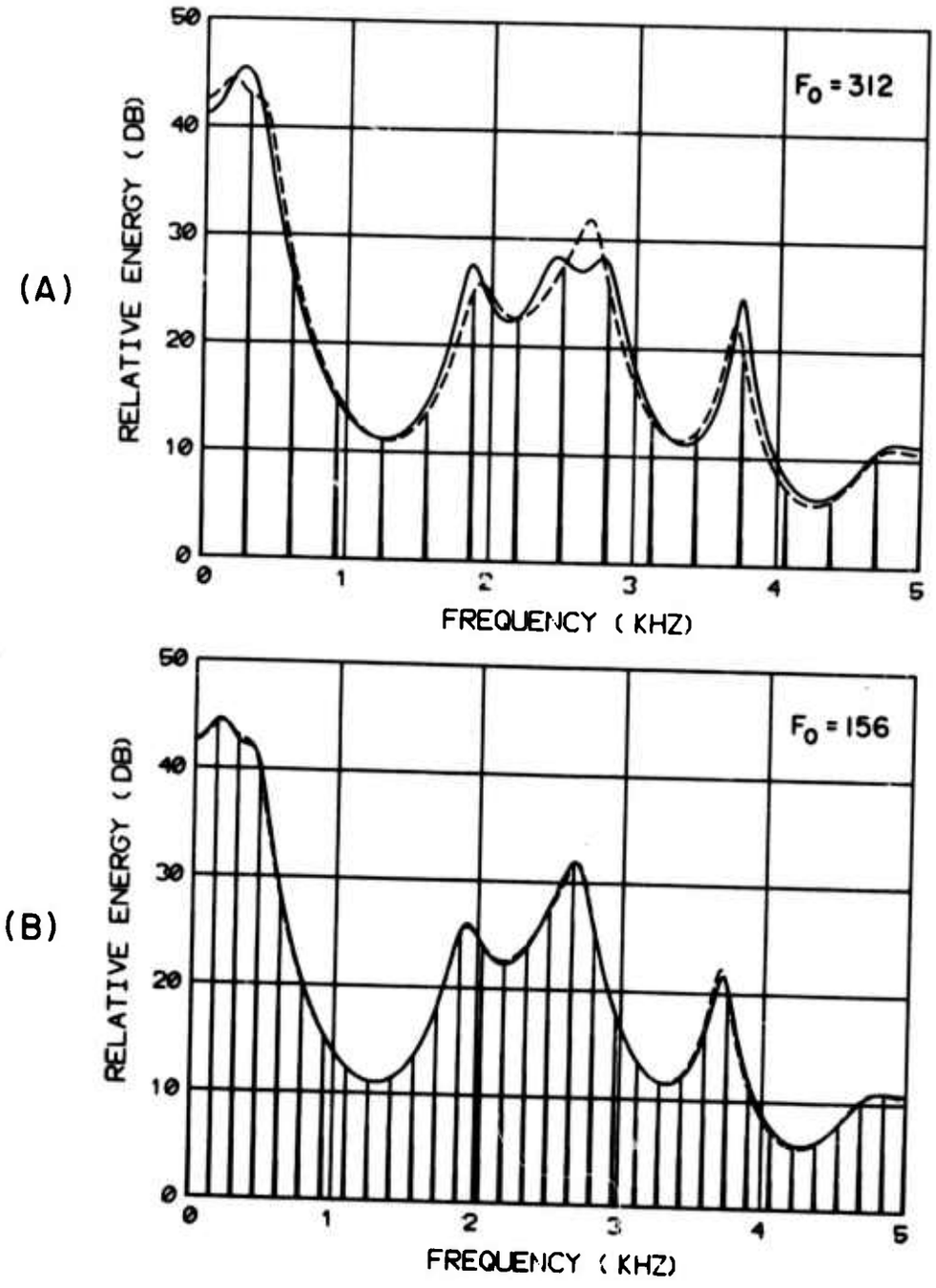


Fig. 8. LP modeling of harmonic spectra. (See Fig. 7)

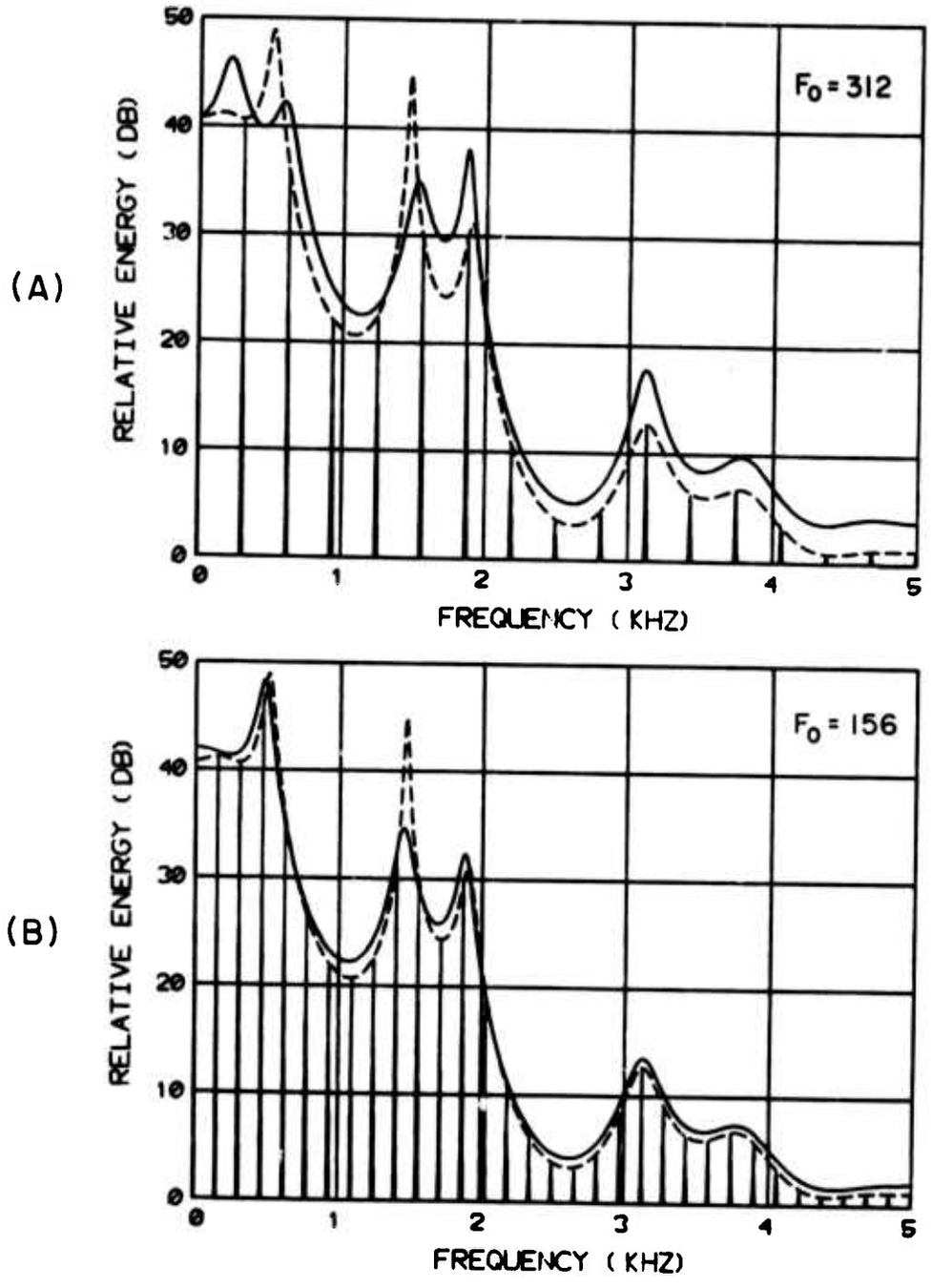


Fig. 9. LP modeling of harmonic spectra. (See Fig. 7)

true only if the period $\tau \geq 2p_0$ samples.) It is unfortunate that the all-pole spectrum resulting from LP modeling does not yield the spectrum we desire.

Another relevant spectrum is that of a single pitch period τ; let that be $Q(\omega)$. It is well known that $Q(\omega)$ is an all-zero spectrum that coincides with $P_0(\omega)$ only at the harmonics $n\omega_0$ i.e. $Q(n\omega_0) = P_1(n\omega_0) = P_0(n\omega_0)$. However, since $Q(\omega)$ is otherwise not equal to $P_0(\omega)$, applying LP modeling to $Q(\omega)$ with $p=p_0$ will result in an LP spectrum $\hat{Q}(\omega)$ that is still different from the all-pole $P_0(\omega)$ and also different from the LP spectrum $\hat{P}_1(\omega)$ corresponding to the discrete spectrum $P_1(\omega)$, i.e. $\hat{Q}(\omega) \neq \hat{P}_1(\omega) \neq P_0(\omega)$.

It would seem from the above that LP analysis of periodic signals (especially those with high fundamental) is doomed to be of a very approximate nature. Indeed, if nothing is known about the transfer function of the system, there is a basic loss of information in the spectrum of the periodic signal that is irrecoverable. This is true whether one uses linear prediction or some other form of analysis. However, the previous discussion shows that even when we are given the extra information that the system transfer function is all-pole, LP analysis does not seem to be able to recover that all-pole spectrum. The reason, of

course, is that nowhere in the analysis did we actually use the fact that $P_o(\omega)$ is all-pole. For example, in computing $\hat{P}_1(\omega)$ from the line spectrum $P_1(\omega)$, we did not make use of the fact that $P_1(n\omega_o) = P_o(n\omega_o)$ and that $P_o(\omega)$ is all-pole. In fact, LP analysis does not allow us to use that information.

All is not lost, however. The trick is to use the fact that $P_1(n\omega_o) = P_o(n\omega_o)$ to generate $P_o(\omega)$ for all ω , and then to apply LP analysis to that, resulting in an LP spectrum identical to $P_o(\omega)$. In order to generate all of $P_o(\omega)$ from $P_1(n\omega_o)$ we use the important fact that the autocorrelation of an all-zero spectrum with p_o zeros is equal to zero for lags $|k|>p_o$. For example, from (26) we see that the autocorrelation b_k of the all-zero inverse filter $A(z)$ is zero for $|k|>p$. Since $P_o(\omega)$ is all-pole, its inverse $P_o^{-1}(\omega)$ is all-zero. Let the autocorrelation of $P_o^{-1}(\omega)$ be r_k. Then $r_k=0$ for $|k|>p_o$,

$$P_o^{-1}(\omega) = \sum_{k=-p_o}^{p_o} r_k \, e^{-jk\omega} \quad , \tag{31}$$

and

$$P_o^{-1}(n\omega_o) = \sum_{k=-p_o}^{p_o} r_k \, e^{-jkn\omega_o} \, . \tag{32}$$

But since $P_1(n\omega_o) = P_o(n\omega_o)$ we must have

$$P_1^{-1}(n\omega_o) = P_o^{-1}(n\omega_o) \quad .$$

Therefore,
$$P_1^{-1}(n\omega_o) = \sum_{k=-p_o}^{p_o} r_k\, e^{-jkn\omega_o}$$

$$= \sum_{k=-p_o}^{p_o} r_k\, e^{-j2\pi kn/\tau} \quad , \quad 0 \le n \le \tau-1 \ , \qquad (33)$$

where τ is the number of samples in a pitch period. If we define

$$\tilde{r}_k = r_k \ , \quad 0 \le k \le \left[\frac{\tau}{2}\right] \ ,$$

$$\text{and} \quad \tilde{r}_{\tau-k} = \tilde{r}_k \ , \qquad (34)$$

then
$$P_1^{-1}(n\omega_o) = \sum_{k=0}^{\tau-1} \tilde{r}_k\, e^{-j2\pi kn/\tau} \quad , \quad 0 \le n \le \tau-1 \ . \qquad (35)$$

But (35) is a τ-point DFT, whose inverse is given by

$$\tilde{r}_k = \frac{1}{\tau} \sum_{n=0}^{\tau-1} P_1^{-1}(n\omega_o)\, e^{-j2\pi kn/\tau} \quad , \quad 0 \le k \le \tau-1. \qquad (36)$$

Therefore, from (36), (34) and (31), one can reconstruct $P_o(\omega)$. This is done as follows:

1. Compute the inverse of the line spectrum:
   $P_1^{-1}(n\omega_o) = 1/P_1(n\omega_o)$, $0 \le n \le \tau-1$.
2. Compute the inverse DFT of $P_1^{-1}(n\omega_o)$ using (36).
   With (34), this yields the autocorrelation function r_k.
3. Compute the all-zero spectrum $P_o^{-1}(\omega)$ from (31) for a large number of frequencies.

(37)

4. Compute $P_o(\omega) = 1/P_o^{-1}(\omega)$.

If the spectrum $Q(\omega)$ of a single pitch period is given, then the first thing to do is to sample $Q(\omega)$ at the harmonics. This yields the line spectrum $P_1(n\omega_0)$. Then follow the procedure (37) above to compute $P_0(\omega)$. Applying LP analysis to $P_0(\omega)$ with $p=p_0$ will then yield an LP spectrum equal to $P_0(\omega)$.

Above we have shown how to recover the complete all-pole spectrum given a finite number of equally spaced points on it. The only restriction is that the number of harmonics in the spectrum be at least equal to the number of poles. This can be thought of as a method of "smoothing" the discrete spectrum. The smoothing is done by resorting to the autocorrelation of the inverse spectrum. Thus we might label this type of smoothing as <u>inverse autocorrelation smoothing</u>. Because this method of smoothing is based on an all-pole assumption for the spectrum, its application to more general cases has anticipated problems. As a simple example, let us assume that the given harmonic spectrum is all-pole but noisy (e.g. as a result of quantization). This case has arisen in our experiments in speech compression [15] where selected spectral values are used as transmission parameters. We employ the procedure given in (37) above to recover the linear prediction coefficients. Problems arise upon quantization of the spectral values to less than 5 bits. The autocorrelation coefficients as computed from (36) lose their positive definiteness

which results in a smoothed spectrum that is negative in certain regions. This, in turn, results in an unstable linear prediction filter with some poles outside the unit circle. There are ways to remedy these situations in a reasonable manner [15], but the message is clear that one should anticipate such problems. The same problems arise if the original spectrum contains zeros as well as poles. It should be emphasized, however, that these problems arise when the number of harmonics in the spectrum is small, i.e. on the order of the number of poles. If the number of harmonics is at least twice the number of poles the problems are not likely to arise. However, for those cases, regular LP analysis on the line spectrum produces satisfactory results, thus obviating the need to use the procedure in (37).

VI. LINEAR PREDICTION VS. ANALYSIS-BY-SYNTHESIS

An important aspect of any fitting or matching procedure is the properties of the error measure that is employed, and whether those properties are commensurate with certain objectives. In the spectral analysis of speech, a common objective is to have the model spectrum $\hat{P}(\omega)$ approximate the envelope of the signal power spectrum $P(\omega)$. In this section we shall explore in some detail the properties of the error measure used in LP analysis and then compare it to the error measure used in AbS, always using as our criterion of goodness the ability of each matching procedure to approximate the envelope of the signal spectrum.

LP Error Measure

One important consideration in estimating the spectral envelope is the determination of an optimal value for p, the number of poles in the model spectrum. This topic has been discussed elsewhere [8,9] and we shall not pursue it in this paper. However, assuming that somehow we know this optimal value of p, there remains the question of whether minimization of the error measure in (5) will result in a good estimate of the spectral envelope.

For each value of p, minimization of the error measure E in (5) leads to the minimum error E_p in (11). It can be shown

[8] that E_p is also equal to

$$E_p = e^{\hat{c}_0} , \tag{38}$$

where

$$\hat{c}_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \hat{P}(\omega)\, d\omega \tag{39}$$

is the zeroth coefficient (quefrency) of the cepstrum corresponding to $\hat{P}(\omega)$. E_p can also be interpreted as the geometric mean of the model spectrum $\hat{P}(\omega)$. E_p decreases monotonically as p increases [8], and the minimum occurs as $p \to \infty$, where $\hat{P}(\omega)$ becomes identical to $P(\omega)$, and (38) reduces to

$$E_{min} = E_\infty = e^{c_0} , \tag{40}$$

where c_0 is obtained by subsituting $P(\omega)$ for $\hat{P}(\omega)$ in (39). If $P(\omega)$ is a p_0-pole spectrum then $E_p = E_{min}$ for all $p \geq p_0$. The absolute minimum error is a function of $P(\omega)$ only, and is equal to its geometric mean, which is always positive and usually nonzero for speech spectra. This is a curious result, because it says that the minimum error can be nonzero even when the matching spectrum $\hat{P}(\omega)$ is identical to the matched spectrum $P(\omega)$. This unusual property is due to the fact that the error measure in (5) is defined as the average of the ratio of two quantities and not their difference as is usual with most error measures such as the mean squared error.

Let the ratio of $P(\omega)$ to $\hat{P}(\omega)$ be given by

$$E(\omega) = \frac{P(\omega)}{\hat{P}(\omega)} \quad . \tag{41}$$

Then from (23) we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} E(\omega) d\omega = 1 \ , \text{ for all p.} \tag{42}$$

$E(\omega)$ can be interpreted as the "instantaneous error" between $P(\omega)$ and $\hat{P}(\omega)$ at frequency ω. Equation (42) says that the arithmetic mean of $E(\omega)$ is equal to 1, which means that there are values of $E(\omega)$ greater and less than 1 such that the average is equal to 1. (Except for the special case when $P(\omega)$ is all-pole, the condition $E(\omega)=1$ for all ω is true only as $p \to \infty$.) In terms of the two spectra, this means that $P(\omega)$ will be greater than $\hat{P}(\omega)$ in some regions and less in others such that (42) applies. However, the contribution to the total error is more significant when $P(\omega)$ is greater than $\hat{P}(\omega)$ than when $P(\omega)$ is smaller, e.g. a ratio $E(\omega)=2$ (+3dB) contributes more to the total error than a ratio of 1/2(-3dB). We conclude that, after the minimization of error, we expect a better fit of $\hat{P}(\omega)$ to $P(\omega)$ where $P(\omega)$ is greater than $\hat{P}(\omega)$, than where $P(\omega)$ is smaller. For example, if $P(\omega)$ is the power spectrum of a quasi-periodic signal (such as a sonorant), then most of the energy in $P(\omega)$ will

exist in the harmonics, and very little energy will reside between harmonics. The error measure in (5) insures that the approximation of $\hat{P}(\omega)$ to $P(\omega)$ is far superior at the harmonics where the energy is greater, than between the harmonics where there is very little energy. Since $\hat{P}(\omega)$ is expected to be a smooth spectrum (this is insured by choosing an appropriate value of p), we conclude that minimization of the error measure in (5) results in a model spectrum $\hat{P}(\omega)$ that is a good estimate of the spectral envelope of the signal spectrum $P(\omega)$. It should be clear from the above that the importance of the goodness of the error measure is not as crucial when the variations of the signal spectrum from the spectral envelope are much less pronounced, such as spectra of unvoiced stops, spectra of single pitch periods, and ordinary filter-bank spectra.

Another important property of this estimation procedure is that, because the contributions to the total error are determined by the ratio of the two spectra, the matching process should perform uniformly over the frequency range of interest, irrespective of the shaping of the speech spectral envelope.

The error measure E is similar in its properties to an error measure used by Itakura and Saito [12,13] in their maximum likelihood method which results in the same set of equations (10). Their error measure is also "more sensitive to the

spectral peaks and less to the dips" [12]. They conclude that for the purposes of synthesis this is a good property because the ear is more sensitive to peaks than to dips in the spectrum. Itakura and Saito were not explicit in what they meant by spectral peaks and dips. There are two likely interpretations: (1) The peaks correspond to harmonic peaks, and the dips are those between the harmonic peaks. (2) The peaks correspond to formants and the dips are the valleys in between. The second interpretation is the one Flanagan [14] gives in his review of Itakura and Saito's work. Flanagan states that "the minimization results in a fit which is more sensitive at the spectral peaks than in the valleys between the formants" [14]. We believe both interpretations to be correct, but under very different conditions. It all depends on the number of poles in the model spectrum. If the number of poles is less than the necessary number to characterize all the formants in the spectrum then indeed the fit could be better at the formant peaks than in the valleys. On the other hand, if the number of poles is greater than or equal to the minimum number of poles necessary to represent the spectral envelope as in Figs. 1 and 5a, then the fit in the valleys between the formants is just as good as the fit at the formant peaks. In this case, the first interpretation given above is more appropriate. Indeed, it is a fundamental property of the error measure E in (5) that given any

peaks and dips one wishes to define, one can always find a value for p, the number of poles, such that the fit is equally good at the peaks and dips. In fact, we know from (22) above that as $p \to \infty$, the model spectrum fits the signal spectrum exactly, all peaks and dips included. This, of course, is also true for all $p \geq p_0$ if $P(\omega)$ is a p_0-pole spectrum.

It is clear from the above that the number of model spectrum poles plays a crucial role in determining how the model spectrum fits the signal spectrum. Since interpretations in terms of peaks and dips can be misleading if not stated carefully, we prefer to interpret the matching process by the relation of the values of the signal spectrum $P(\omega)$ relative to those of the model spectrum $\hat{P}(\omega)$. We merely state that, after error minimization, the fit will be better for values of $P(\omega) > \hat{P}(\omega)$ than for values of $P(\omega) < \hat{P}(\omega)$. For spectral envelope estimation with an appropriate number of poles, this guarantees us that harmonic peaks $(P(\omega) > \hat{P}(\omega))$ are matched better than the dips in between $(P(\omega) < \hat{P}(\omega))$, resulting in a good spectral envelope match. For purposes of synthesis, a better spectral envelope fit results in better synthesis, i.e. a better "perceptual fit".

Comparison With AbS

In AbS [3] the error measure that was normally used is given

(in our notation) by:

$$E' = \int_{\omega} E'(\omega)\,d\omega \tag{43}$$

where

$$\begin{aligned} E'(\omega) &= [\log P(\omega) - \log \hat{P}(\omega)]^2 \\ &= \log\left[\frac{P(\omega)}{\hat{P}(\omega)}\right]^2 \\ &= [\log E(\omega)]^2 \quad . \end{aligned} \tag{44}$$

Here $\hat{P}(\omega)$ is the model spectrum, $E(\omega)$ is the ratio of the two spectra as in (41), and the integration in (43) is over the frequency range of interest. Minimizing E' is equivalent to minimizing the mean squared error between the two log spectra. In contrast to the error measure E in LP, here a minimum error of zero is possible, namely when the two spectra are identical.

The error measures E and E' in (5) and (43) are similar in that the contributions to the total error are functions of the ratio of the two spectra. We have already mentioned that this fact makes the matching process perform uniformly over the frequency range of interest. However, the error measure E in LP spectral matching has two advantages over E': (a) For an all-pole model spectrum, the minimization of E in (5) leads to

a solution where the coefficients of the resulting $\hat{P}(\omega)$ are computed simply by solving a set of simultaneous linear equations, while the minimization of E' has to be done iteratively. (b) For many cases of interest, E is a superior error measure to E' if a spectral envelope is desired. This is clear if one notes from (44) that contributions to the total error E' are made equally whether $P(\omega)>\hat{P}(\omega)$ or $P(\omega)<\hat{P}(\omega)$, e.g. a ratio $E(\omega)=2$ (+3dB) contributes equally to the total error E' as a ratio of 1/2 (-3dB). This means that energy at the harmonics (in voiced sounds) and the lack of energy between harmonics contribute equally to the total error. This, of course, will not lead to a good spectral envelope. One can dramatize the difference between the error measures E and E' by assuming that the signal spectrum $P(\omega)=0$ for some range of frequencies (no matter how small). The ratio $E(\omega)$ will be zero for the same range, but $E'(\omega)$ in (44) will be infinite. The effect of this range of frequencies on the total error is nil for E and total for E'. It is clear that for cases where the variations of the signal spectrum about the spectral envelope are large, E is a preferable measure of error to E'.

But then, traditional AbS methods have generally used already smoothed spectra, in which case it is not exactly clear which error measure is to be preferred. For the special case when the signal spectrum is all-pole we know that both LP and

AbS error minimization result in a model spectrum that is identical to the signal spectrum. (A salient difference, though, is that the minimum error E' in AbS will be zero.) For other smooth signal spectra there is independent evidence [15] that the AbS error measure might result in a better spectral fit. However, for FFT-generated spectra (from a time signal) we believe that linear prediction will generally be superior to AbS.

Comparison for Discrete Spectra

Another point of comparison between LP and AbS is in the case of discrete spectra. This case is of particular interest because AbS techniques were largely applied to filter bank spectra. We shall consider only two types of spectra - harmonic spectra and filter bank spectra. Both types of spectra will be considered to be samples on a smooth spectral envelope.

The definition of error for AbS is obtained by replacing the integral in (43) by a summation

$$E' = \sum_{n=0}^{N-1} \log\left[\frac{P(\omega_n)}{\hat{P}(\omega_n)}\right]^2 \quad . \tag{45}$$

The comparison now is between E' in (45) and E in (28). The absence of the factor G^2/N in (45) is irrelevant to this discussion.

An example which will put the issues into focus is that given in Section V, where the signal spectrum is an all-pole harmonic spectrum $P_1(\omega)$ as defined by (30), i.e. the harmonics lie on a p_0-pole spectrum $P_0(\omega)$. We have seen that LP analysis will not result in the desired envelope spectrum $P_0(\omega)$, as was illustrated in Figs. 7-9. On the other hand, one can show that by minimizing E' in (45) with $P(\omega_n) = P_1(n\omega_0)$, the model AbS spectrum will be identical to $P_0(\omega)$ for $p=p_0$. (The only possible restriction is that the number of harmonics be at least equal to the number of poles.) This is clear by noting that the absolute minimum value that E' in (45) can have is zero, and this occurs only when the two spectra are equal at each frequency ω_n. Since in this example we know that there is a unique all-pole spectrum $P_0(\omega)$ that is equal to $P_1(\omega)$ at each frequency $\omega_n=n\omega_0$, we conclude that the all-pole model spectrum $\hat{P}_1(\omega)$ will result in an error $E'=0$, and therefore must be identical to $P_0(\omega)$.

The above example shows that for modeling of all-pole harmonic spectra, AbS is clearly superior to LP. One could argue that for this special case of all-pole harmonic spectra, it is possible to use "inverse autocorrelation smoothing" as described in Section V to recover the all-pole spectrum so that LP analysis will result in the desired spectrum. However, as we pointed out earlier, this method of smoothing is sensitive to spectral noise

and to the existence of zeros in the signal spectrum; its use is generally not recommended. We do not mean to imply in the above arguments that LP analysis should not be used at all with harmonic spectra. We merely point out that AbS gives better results, but at a much higher computational cost. If the results shown in Figs. 7-9 are satisfactory for the application one has in mind, then clearly LP analysis is to be preferred because of the lower cost. If more accurate results are desired then one must pay the price inherent in AbS. The same comments also apply to modeling of filter bank spectra.

The reader might sense a contradiction between the above conclusions and those made earlier in this section. (i) Earlier we stated that, especially for the case of spectra of voiced sounds where the energy is mainly concentrated around the harmonics, such as in Fig. 5a, LP analysis is superior to AbS in that it results in a better spectral envelope fit. (ii) On the other hand, we have shown above that for the case of harmonic spectra, such in Figs. 7-9, AbS is superior to LP. The contradiction is only apparent. The two types of harmonic spectra mentioned above are radically different in the way they affect error minimization. The signal spectrum in Fig. 5a makes large excursions from the spectral envelope. While these excursions are of little importance in LP error minimization, they are

disastrous to AbS error minimization. In contrast, in Figs. 7-9, only the values at the harmonics are included in the error, so that there are no large excursions to upset AbS error minimization. It is not that LP does better in case (i), e.g. Fig. 5a, it is that AbS does much worse. In fact, LP performs about the same in cases (i) and (ii). The conclusions concerning LP analysis as depicted in Figs. 7-9 also apply to the case in Fig. 5a. The problem is that if one has to deal with case (i) then AbS does not perform well and there is little choice but to use LP analysis. An interesting solution to this problem is to convert case (i) to case (ii) and then apply AbS instead of LP. This can be done in Fig. 5a, for example, by "peak picking" the harmonics, i.e. retain the values only at the harmonic peaks and discard all other values, then apply AbS to the resulting line spectrum. That should give better results than straight LP, especially for high fundamentals. Another possibility is to take the spectrum of a single pitch period, sample it at the harmonics and then use AbS. The main obstacle, however, is the computational cost associated with AbS. The attraction of LP modeling is its simplicity; the price that one pays is that the model spectrum can have only poles, and a degradation in performance is expected with an increase in pitch frequency.

VII. ALL-ZERO MODELING

We have seen in section II that if the model spectrum is all-pole then the minimization of the LP error in (5) leads to a set of linear equations (10) which can be easily solved for the parameters of the model. It is straightforward to show that if the model spectrum contains zeros (with or without poles), then the minimization of (5) leads to a set of nonlinear equations whose solution is generally iterative and not always readily convergent. Computation-wise then, LP analysis that includes zeros in the model offers no distinct advantages over AbS.

However, if the model spectrum is all-zero then the problem can be reformulated such that a suboptimal solution can be obtained noniteratively. The idea is quite simple: Invert the signal spectrum and apply an all-pole LP analysis, then invert the all-pole LP spectrum to obtain the desired all-zero model. We shall call this process inverse LP modeling . This solution is clearly reasonable, and on the surface even seems to be optimal. Unfortunately, there is a problem. Below we discuss this problem and show how to deal with it.

We state again that our purpose in spectral modeling is to obtain a good fit to the envelope of the signal spectrum. The problem in the solution given above is that, in general, the envelope of the inverted spectrum is not equal to the inverse

envelope of the spectrum. For example, if we invert the signal spectrum in Fig. 5a, then the harmonic peaks become valleys and the valleys between the harmonics become the new peaks. We know that LP analysis on this inverted spectrum will follow these new peaks whose envelope is not the one we are after. This problem is not so severe if the signal spectrum is smooth relative to the order of the model. For example, if the signal spectrum consists of q zeros only, then the above method leads to the correct solution for p=q. Therefore, the solution to our problem is to smooth the signal spectrum before we apply inverse LP analysis. However, smoothing introduces a certain amount of error. Therefore, inverse LP modeling on the smoothed spectrum is only a suboptimal solution. The type and degree of smoothing can effect the final result appreciably. Below we discuss these matters briefly.

The degree to which smoothing is performed must depend on the order of the model considered. For example, a large amount of smoothing can be tolerated if the order of the model is small. In general, the simplest and perhaps most effective way to determine the degree of smoothing is by inspection of the results.

There are several types or methods of spectral smoothing. One can apply a low pass filter to the spectrum (autocorrelation smoothing) or to the log spectrum (cepstral smoothing).

Autocorrelation smoothing has been used extensively by statisticians. Cepstral smoothing is a more recent development that has been employed in speech and picture processing. Another method of smoothing that has become quite popular recently is LP smoothing. Indeed, LP modeling can be thought of as just another method of smoothing the spectrum. The degree of smoothing is controlled by the order of the predictor. Usually, the order of the predictor p is chosen to be much larger than the number of zeros in the model q. In this method, the whole procedure is as follows: (a) Perform a regular p pole LP analysis on the signal spectrum, where p>>q. (b) Compute the corresponding LP spectrum and invert it. (c) Perform a q-pole LP analysis on the inverted spectrum. The resulting predictor coefficients are the desired parameters of the all-zero model.

We point out that in speech analysis all-zero modeling can be used to study the spectral characteristics of glottal pulses.

VIII. CONCLUSIONS

Linear predictive analysis was presented as a problem in spectral modeling in which the signal spectrum is modeled by an all-pole spectrum through the minimization of an error measure given by the integrated ratio of the signal and model spectra. The parameters of the all-pole model are obtained as the solution of a set of linear equations. The only values needed for the computation of all p parameters are the first p+1 autocorrelation coefficients which are computed from the signal spectrum by a simple Fourier transform. Alternatively, the autocorrelation coefficients can be computed from the time signal, if available.

The spectral formulation leads to the method of selective linear prediction where selected portions of a spectrum can be fitted by an all-pole spectrum. This method allows for arbitrary spectral shaping in the frequency domain, thus obviating the need for any special time domain filtering. In addition, different portions of a spectrum can be fitted by different numbers of poles, a property that is useful in speech recognition applications. The method is also applicable to linear predictive speech compression systems where different sampling rates can be simulated without the need for sharp filtering or down sampling.

LP analysis has also been applied to the modeling of discrete spectra, such as harmonic and filter bank spectra. It was shown that the modeling process has definite problems as the number of spectral lines decreases, i.e. as the fundamental frequency increases. This has clear implications for the analysis of high-pitched voices, such as female and children speech. For the special case when the harmonic spectrum is a sampled all-pole spectrum, we were able to recover the all-pole spectrum by first applying inverse autocorrelation smoothing. However, this method of smoothing was not recommended as a general method of dealing with the problems associated with high fundamentals.

A detailed comparison was given between LP modeling and analysis-by-synthesis (AbS) in which the error measure is defined as the average of the square of the difference between the signal and model log spectra. The two methods were seen to have two properties in common: (a) The spectral matching can be done selectively to any portion of the spectrum, and (b) both error criteria are functions of the ratio of the original and model spectra, which results in a matching process that performs uniformly over the frequency range of interest. For the special case of an all-pole model, LP analysis was seen to offer two important advantages: (a) The computations for the spectral parameters are straightforward and noniterative, and (b) if the

time signal is available there is no need to compute the spectrum first. However, the major difference between LP and AbS modeling is in the quality of match between the model and signal spectra. If the variations of the signal spectrum about the model spectrum are large, then LP analysis is preferable to AbS. This is usually the case if the signal spectrum is FFT-derived from a time signal. However, if the signal spectrum is smooth relative to the model spectrum, then AbS is expected to give better results than LP analysis. This occurs with filter bank spectra and cepstrally (or otherwise) smoothed spectra.

Finally, we gave a suboptimal solution to the problem of all-zero modeling using LP analysis. The solution is simply to apply all-pole LP modeling to the inverted spectrum. This, however, requires that the spectrum be smoothed before inversion.

REFERENCES

[1] Koenig, W., H.K. Dunn, and L.Y. Lacey, "The Sound Spectrograph," J. Acoust. Soc. Am., vol. 18, 19-49, 1946.

[2] Fant, G., Acoustic Theory of Speech Production, Mouton & Co., 's-Gravenhage, The Netherlands, 1960.

[3] Bell, C.G., H. Fujisaki, J.M. Heinz, K.N. Stevens, and A.S. House, "Reduction of Speech Spectra by Analysis-by-Synthesis Techniques," J. Acoust. Soc. Am., vol. 33, No. 12, 1725-1736, Dec. 1961.

[4] Paul, A.P., A.S. House, and K.N. Stevens, "Automatic Reduction of Vowel Spectra: An Analysis-by-Synthesis Method and its Evaluation," J. Acoust. Soc. Am., vol. 36, No. 2, 303-308, Feb. 1964.

[5] Fujimura, O., "Analysis of Nasal Consonants," J. Acoust. Soc. Am., vol. 34, No. 12, 1865-1875, Dec. 1962.

[6] Mathews, M.V., J.E. Miller, and E.E. David, "Pitch Synchronous Analysis of Voiced Sounds," J. Acoust. Soc. Am., vol. 33, 179-186, 1961.

[7] Olive, J.P., "Automatic Formant Tracking by a Newton-Raphson Technique," J. Acoust. Soc. Am., vol. 50, 661-670, 1971.

[8] Makhoul, J. and J. Wolf, "Linear Prediction and the Spectral Analysis of Speech," BBN Report No. 2304, Bolt Beranek and Newman Inc., Cambridge, Mass., August 1972.

[9] Makhoul, J., "Spectral Analysis of Speech by Linear Prediction," IEEE Trans. Audio and Electroacoustics, vol. AU-21, pp. 140-148, June 1973.

[10] Grenander, U., and G. Szegö, Toeplitz Forms and their Applications, Univ. of California Press, Berkeley, 1958.

[11] Atal, B.S., "Influence of pitch on formant frequencies and bandwidths obtained by linear prediction analysis," J. Acoust. Soc. Am., vol. 55, Supplement, S81, Spring 1974.

[12] Itakura, F. and S. Saito, "Analysis Synthesis Telephony based on the Maximum Likelihood Method," Paper C-5-5, Proc. of the 6th International Congress on Acoustics, Tokyo, Japan, Aug. 21-28, 1968.

[13] Itakura, F., and S. Saito, "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies," Electronics and Comm. in Japan, vol. 53-A, No. 1, 36-43, 1970.

[14] Flanagan, J.L., Speech Analysis Synthesis and Perception, Academic Press Inc., New York, 1965, Second Edition 1972.

[15] Makhoul, J. and R. Viswanathan, "Quantization Properties of Transmission Parameters in Linear Predictive Systems," BBN Report No. 2800, Bolt Beranek and Newman Inc., Cambridge, Mass., April 1974.

[16] Markel, J.D. and A.H. Gray, Jr.,""On Autocorrelation Equations with Application to Speech Analysis," IEEE Trans. Audio Electroacoust., vol. AU-21, No. 2, 69-79, April 1973.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Mass. 02138

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP
AD-A108349

3. REPORT TITLE
"Selective Linear Prediction and Analysis-by-Synthesis in Speech Analysis"

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Technical Report

5. AUTHOR(S) *(First name, middle initial, last name)*
John I. Makhoul

6. REPORT DATE
April 1974

7a. TOTAL NO. OF PAGES
58

7b. NO. OF REFS
16

8a. CONTRACT OR GRANT NO.
DAHC-71-C-0088

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
BBN Report No. 2578
A.I. Report No. 13

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

Linear prediction is presented as a spectral modeling technique in which the signal spectrum is modeled by an all-pole spectrum. The method allows for arbitrary spectral shaping in the frequency domain, and for modeling of continuous as well as discrete spectra (such as filter bank spectra). In addition, using the method of selective linear prediction, all-pole modeling is applied to selected portions of the spectrum, with applications to speech recognition and speech compression. Linear prediction is compared with traditional analysis-by-synthesis techniques for spectral modeling. It is found that linear prediction offers computational advantages over analysis-by-synthesis, as well as better modeling properties if the variations of the signal spectrum from the desired spectral model are large. For relatively smooth spectra and for filter bank spectra, analysis-by-synthesis is judged to give better results. Finally, a suboptimal solution to the problem of all-zero modeling using linear prediction is given.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

Unclassified
Security Classification

A-31408

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Linear Prediction						
Selective Linear Prediction						
Spectral Analysis						
All-Pole Models						
All-Zero Models						
Autoregressive Models						
Moving Average Models						
Speech Analysis						
Speech Recognition						
Speech Compression						
Analysis-by-Synthesis						
Signal Processing						